



**UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION IX
75 Hawthorne Street
San Francisco, CA 94105**

SENT VIA EMAIL AS PDF

April 17, 2020

Elizabeth Roddy
Remedial Project Manager
Department of the Navy
Base Realignment and Closure Program Management Office West
33000 Nixie Way, Building 50
San Diego, CA 92147

Subject: EPA Comments on the Draft Background Soil Study Report
Hunters Point Naval Shipyard Superfund Site

Dear Ms. Roddy:

Thank you for providing the "Draft Background Soil Study Report" for the Hunters Point Naval Shipyard Superfund Site in San Francisco, California (Site). The draft report was prepared by CH2M Hill for the Department of the Navy and is dated February 2020. In the report, the Navy summarizes an evaluation of radiological soil sampling data from five "reference background areas" (RBAs) believed to be unaffected by past Navy activities at the Site. The Navy collected soil samples in August and September 2019, under EPA oversight.

The results of this report will guide the Navy's planned retesting and, if needed, remediation of areas where the previously collected radiological data are unreliable. In the report, the Navy uses the new data to calculate site-specific "background threshold values (BTVs)" for six radionuclides of concern at the Site. BTVs specify the concentrations that we expect to see in the environment in the absence of Superfund site contamination. These values are important because Superfund site cleanups generally do not clean up below background levels.

Developing appropriate BTVs for the Site is complicated by the natural occurrence of some of the radionuclides of concern, global fallout from atomic bomb testing in the 1950s and 60s, and the relatively large range in background radionuclide concentrations in Site soils. The large range is reflected in the 2019 sampling results and is in part due to the complex fill history of the Site. Much of the Site is made of rock and soil cut from nearby hills, sediments dredged from San Francisco Bay, and other offsite sources of fill material.

In the report, the Navy presents five BTVs for each radionuclide evaluated (one BTV for each of the five RBAs). We understand that the Navy's preferred approach is to use the highest of the five BTVs sitewide. We are concerned that this approach would overestimate the background radionuclide

concentrations in an unknown but potentially significant fraction of the soils to be retested. We describe our concerns with this approach in comment #19 in an enclosure to this letter.

We propose, as an alternative approach, calculating BTVs based only on sampling results from the offsite San Bruno Mountain location. This approach would decrease the chances of concluding that a sample result reflects background when it actually results from Site contamination (i.e., a “false negative” error). A drawback of this approach is that it would increase the chances of concluding that a sample result reflects Site contamination when it actually reflects background (i.e. a “false positive” error). Comment #19 describes the proposed approach in more detail.

Remediation goals (i.e., cleanup levels) at Superfund sites are generally not set below background levels. At the Site, any BTV determined to be above the current remediation goal is expected to become the new remediation goal. The Navy’s preferred approach would establish BTVs above the current remediation goals for three radionuclides. EPA’s proposed approach, using sampling results from the offsite San Bruno Mountain location, would have a more limited impact. It would establish the BTV for one radionuclide (cesium-137) above the current remediation goal. The cesium-137 remediation goal could therefore increase from 0.113 picoCuries per gram (pCi/g) to 0.141 pCi/g. A cesium-137 remediation goal of 0.141 pCi/g would remain protective of human health, falling within the EPA cancer risk range of 10⁻⁶ to 10⁻⁴.

If a change in the remediation goal is appropriate, the Navy would need to comply with the post-ROD change process outlined in EPA Superfund guidance.

Whichever approach is used to develop and apply the BTVs, there will be a risk of both false positive and false negative errors when Site data are compared to the BTVs. We believe that the use of the San Bruno Mountain data better balances the relative risk of the two types of error compared to the Navy’s preferred approach.

Please see our enclosed comments.

Please contact me at 415-972-3181 or praskins.wayne@epa.gov with any questions.

Sincerely,

WAYNE
PRASKINS

Digitally signed by
WAYNE PRASKINS
Date: 2020.04.17
11:52:57 -07'00'

Wayne Praskins
EPA Project Manager

cc: Nina Bacey, California Department of Toxic Substances Control
Jeff Guillory, California Department of Public Health
Shane Reese, California Department of Public Health
Tina Low, San Francisco Regional Water Quality Control Board
Amy Brownell, San Francisco Department of Public Health

Enclosure

**EPA Comments on the Draft Background Soil Study Report
Hunters Point Naval Shipyard Superfund Site
Draft Report dated February 2020; EPA Comments dated April 17, 2020**

#1.	Section 2.4, Soil Sampling (and Table 2-2, Radiological Sample Details)	<p>The table indicates that EPA collected a soil sample to compare to the corresponding Navy sample (a “split sample”) from soil boring 21 in RBA-2. The table does not indicate that a split sample was collected from soil boring 25.</p> <p>EPA collected a total of 30 split samples. At RBA-2, EPA collected a split sample from soil boring 25 (sample HPRBA2-SB25-0102-0919), but not from soil boring 21. Please revise the table to designate the correct split sample location.</p>
#2.	Section 3.1, Onsite Geology, Sections 3.1.1 through 3.1.4, Reference Background Areas	<p>These sections describe soil and material types observed in the borings. Were efforts made to determine whether there were common soil or material types across RBAs and whether they represent distinct radiological statistical populations?</p> <p>Some samples are described as possible road base, durable cover base, ballast, or crushed bedrock. Are these materials distinct radiologically? Do the seven RBA-4 samples determined to be road base contain elevated Cs-137 concentrations, a possibility discussed in Section 2.7?</p>
#3.	Section 3.1, Onsite Geology	There appears to be a typo in the first paragraph: “foo” should be “foot.”
#4.	Section 5.1.3, Data Validation Findings	<p>The text states that “For the evaluation of precision between the native sample and its associated field duplicate, the sample results must be greater than 5 times the minimum detectable concentration (MDC) for the RPD [relative percent difference] criteria to apply. When either the sample or field duplicate results are less than 5 times the MDC, then the RER [relative error ratio] must be less than 1.”</p> <p>Please include the following information (available in the Sampling and Analysis Plan in the Parcel G Work Plan) in the report: formulas/ definitions for RPD and RER, and the RPD criterion.</p> <p>Also, please provide a reference for, or describe the source of, the RER formula.</p>
#5.	Section 5.3, Comparison of Analytical Methods (and Appendix K)	<p>As part of the background study, the Navy used multiple laboratory analytical methods to measure the concentrations of Th-232 and Ra-226. Appendix K graphically compares the results of the analyses.</p> <p>The report states that the Th-232 gamma spectroscopy results appear, on average, greater than results from alpha spectroscopy for both RBA-3 and RBA-S. We agree. We note that one of the graphical approaches, the Q-Q plots, shows that data generated using two of the methods (alpha spectroscopy and gamma spectroscopy) at RBA-S have significantly different slopes (and higher alpha spectroscopy results for samples with</p>

		the highest Th-232 concentrations). Please comment on the significance of this difference.
#6.	Section 5.4, Review of Equilibrium Conditions (and Appendix L)	<p>Section 5.4 provides the results of an evaluation of radionuclides expected to be in secular equilibrium with Ra-226. The report concludes that “the plots show little, if any, secular equilibrium relationships between the 226Ra and the parent radionuclides at the low activity levels observed at HPNS.” This conclusion reflects higher than expected variability between the concentrations of Ra-226 and its parent radionuclides U-238 and Th-230.</p> <p>One possible contributor to the variability is sample heterogeneity, and the use of different sample aliquots for the alpha spectroscopy analyses of the three radionuclides.</p> <p>In circumstances where it is important to minimize sample variability, such as in future evaluations of secular equilibrium, has the Navy considered steps that could be taken to reduce sample heterogeneity? One approach would be to grind and sieve the soil to be used in the alpha spectroscopy analyses to uniform fineness. EPA has used this method with good success and high R-squared values between split samples for analysis of metals in soil.</p> <p>After completion of the draft report, the Navy provided supplemental information on the extent to which the data show secular equilibrium, including the results of additional statistical evaluations. Please include the additional statistical evaluations in the report.</p>
#7.	Section 5.2.4, Reference Background Area-4, (and Table 5-35, RBA-4 - Summary of Combined Analytical Results)	There appears to be one or more errors in the table. For the Th-232 gamma spectroscopy results, the median value (0.857 pCi/g) exceeds the maximum (0.456). Similarly, for Tl-208, the median (0.265) exceeds the maximum (0.108). Please review and correct any errors in the table.
#8.	Section 5.5, USEPA Split Sample Results	<p>The text states that EPA split samples will be compared with the Navy’s sample results using relative percent difference (RPD). EPA is using a different statistic, the duplicate error ratio (DER), to compare split sample pairs.</p> <p>In late April or May, we expect to complete a report summarizing the split sampling effort, including a comparison of the Navy and EPA split sample results.</p>
#9.	Section 6.2.2, Evaluation of Outliers	<p>The reports states that, among other reasons, outliers might be excluded if “...values [were] significantly outside the historical ranges of background data.”</p> <p>Section 1.2 notes that four onsite RBAs were previously used as background areas at the Site. How do results from the 2019 sampling effort compare to historical background data?</p>

#10.	Section 6.2.2, Evaluation of Outliers (and Table 6-1, Statistical Outliers in Combined Surface and Subsurface Depth Intervals)	<p>The table provides the results of the outlier analysis for Ra-226 and Th-232. Please indicate whether the remaining two datasets with detectable radionuclide concentrations (Cs-137 and U-235) were examined for outliers.</p> <p>We note that the 0.477 pCi/g Cs-137 result from RBA-4 is nine or more times greater than the other samples collected in RBA-4 and more than three times greater than the next highest sample from any of the five RBAs. The report states that “Because extreme outliers can adversely affect the results of statistical calculations, it may be advisable at times to remove high-magnitude outliers in background, even if the reasons for these apparently extreme observations are not known.”</p>
#11.	Section 6.2.3, Determination of Soil Groupings	Please clarify why the gamma spectroscopy results were used to evaluate differences in Ra-226 between RBAs and soil depth (rather than alpha spectroscopy or radon emanation) and why the alpha spectroscopy results were used to evaluate Th-232 (rather than gamma spectroscopy).
#12.	Section 6.2.4, Development of Background Threshold Values (and Table 6-5, Summary Statistics for Combined Surface and Subsurface Depth Intervals)	<p>The text states that “Background Threshold Values” (BTVs) were developed for each RBA using a combined dataset representing surface and subsurface soil, and that the statistic chosen for the BTVs is the 95 percent upper simultaneous limit (USL). The USL was chosen “...to control the false positive error rate.”</p> <p>A condition for use of a USL is that the dataset represent a single statistical population. The use of a USL is not advised when the background data set may represent several geological formations or soil types (ProUCL Technical Guide, USEPA, EPA/600/R-07/041, October 2015). As described in Section 6.2.3.2 and summarized in Tables 6-3 and 6-4, three of the five RBAs (RBAs 1, 2, and 4) do not appear to meet this requirement. In each case, the shallow and deep soils appear to represent different statistical populations for the two radionuclides examined (Ra-226 and Th-232). The differences are apparent in the figures comparing the Ra-226 and Th-232 concentrations in shallow and deep soils (Appendix P).</p>
#13.	Section 6.2.4, Development of Background Threshold Values (and Table 6-5, Summary Statistics for Combined Surface and Subsurface Depth Intervals)	<p>Section 6.2.4 states that if a dataset did not include detectable concentrations of a radionuclide (i.e., all non-detects) the highest detection limit of any sample in the dataset was used as the BTV. This was the case for Pu-239 and Sr-90 in all five RBAs and Cs-137 in RBA2. We do not agree with this approach for the Site radiological data because an elevated detection limit can be caused by sampling and analysis limitations that should not dictate the BTV.</p> <p>For example, the proposed BTV for Pu-239 (0.515 pCi/g) is based on the highest reported detection limit in any of the 266 samples collected as part of the background study (sample HPRBAS-SB06-1H02-0919). Detection limits in the other 265 samples were lower, ranging from 0.0333 to 0.498 pCi/g. The higher detection limit in sample HPRBAS-SB06-1H02-0919 appears to be due primarily to small sample size (0.209 grams; less than half the mass of the sample with the lowest detection</p>

		<p>limit). Other factors that can affect the detection limit for a particular sample include the laboratory count time, detector efficiency, and tracer yield.</p> <p>At this time, using a maximum detection limit for Pu-239 or Sr-90 as the basis for a BTV is not expected to have any practical significance on Site cleanup efforts, specifically on the remediation goals (RGs) for the two radionuclides. Neither of the maximum detection limits exceed their corresponding RG (i.e., the maximum detection limit for Pu-239 is 0.515 pCi/g and the RG for Pu-239 is expected to remain 2.59 pCi/g; the maximum detection limit for Sr-90 is 0.150 pCi/g and the RG for Sr-90 is expected to remain 0.331 pCi/g).</p>
#14.	Section 6.2.4, Development of Background Threshold Values (and Table 6-5, Summary Statistics for Combined Surface and Subsurface Depth Intervals)	<p>We understand that the proposed BTV for Uranium-235 is based on the maximum detection limit in RBA2 (0.245 pCi/g, measured in sample HPRBA2-SB21-0304-0919). This concentration is more than twice the highest detected U-235 concentration in RBA2 (0.0899 pCi/g) and much higher than the detection limits for most of the background samples. The elevated detection limit for sample HPRBA2-SB21-0304-0919 appears to be the result of matrix interference. The laboratory report indicates that the sample aliquot used for the laboratory analysis was limited due to the need to reanalyze the sample because of low tracer yield in the initial two analyses.</p> <p>Similar to comment #13, we do not believe that it is appropriate to base a BTV on a sample with an elevated detection limit.</p>
#15.	Section 6.2.4, Development of Background Threshold Values (and Table 6-5, Summary Statistics for Combined Surface and Subsurface Depth Intervals, and Table 6-6, Background Threshold Value Summary for Combined Surface and Subsurface Depth Intervals)	<p>The second to last column in Table 6-5 is labeled "USL95" and the footnote in Table 6-6 refers to all BTVs as USLs. BTVs that are based on Maximum Detections or Maximum Detection Limits are not USLs. Please re-label the columns and modify the footnote. In the table, we suggest including USL95 in the column labeled "Basis" and replacing the column header "USL95" with "BTV".</p>
#16.	Section 7.2, Other Background Data, Section 7.3, HPNS BTVs Comparison to Other Background Values (and Table 7-2,	<p>The table summarizes radionuclide concentrations believed to represent background concentrations in soil in multiple locations in the United States. In the report, the Navy cites eight studies as sources of the data. The report does not present information on soil type or other site-specific parameters that could affect the reported radionuclide concentrations, or the level of confidence that the reported concentrations represent background. Nor does it present information on sampling depth, local</p>

	Background Soil Concentrations of Radionuclides Reported in Literature)	<p>topography, or other factors known to affect the soil concentrations of radionuclides associated with fallout from atomic bomb testing in the 1950s and 1960s, such as cesium and strontium.</p> <p>Section 7.3 includes statements that Site BTVs are representative of background soil because they are within the range reported in the eight studies. The absence of detailed information on the eight studies used to develop the “Range of Literature Values” provided in the table makes it difficult to evaluate this conclusion.</p>
#17.	Section 7.3, HPNS BTVs Comparison to Other Background Values	<p>This section notes that the range of the Site BTVs for Pu-239 is higher than the reported concentrations in the eight studies. It then notes that the detection limits reported for the analyses of RBA samples are consistent with the quantitation limit goals presented in the SAP and are approximately an order of magnitude below the respective RG for the Site. We agree with the latter statements but are unclear on their relevance to the discussion of whether the proposed BTVs are representative of background soil.</p>
#18.	Section 7.3, HPNS BTVs Comparison to Other Background Values	<p>The report includes the following statement: “Based on the uncertainty in the BTV estimates and the potential for false positives..., an exceedance of a BTV for an ROC should not automatically be considered a site release; rather, an exceedance in this case warrants further consideration with respect to literature values and discussion with the project team.”</p> <p>We agree that an exceedance of a BTV for a radionuclide of concern (ROC) does not always indicate a site release (i.e., that there is a chance that a sample result above a BTV could still represent background). The likelihood and frequency with which this may occur depends on the methodology used to determine the BTVs. We do not object to consideration of values in the eight studies but do not believe the values summarized in Table 7.2 necessarily represent background concentrations at the Site. We discuss this further in comment #19.</p>
#19.	Section 6, Statistical Data Evaluation, and Section 7, Use of Background Data	<p>The report provides BTVs for each of six radionuclides at each of the five RBAs. Four of the RBAs are onsite (in areas believed to be uncontaminated by past Navy activities) and one area is offsite (in the San Bruno Mountain State and County Park). The report also provides general statements about planned uses of the BTVs (e.g., “sample data [will be compared] to appropriate RBA data from HPNS”). We are unable, however, to find a clear statement about how the multiple BTVs calculated for each radionuclide will be used at the Site.</p> <p>An approach used at many Superfund sites is to calculate multiple BTVs and match each BTV to specific parcels or portions of a site. We understand that the Navy does not, at this time, believe this approach is practical due to the complex and poorly documented fill history of the Site and past excavation and backfilling in many of the locations to be retested.</p>

This approach remains an option in the future if a subset of the onsite background data can be shown to be representative of a specific parcel or portion of the Site.

A second approach, which we understand the Navy prefers, is to apply the highest BTV calculated for any of the five RBAs presented in the report sitewide. This approach would result in BTVs above the current RGs for three radionuclides. For Cs-137, the BTV would be established at 0.477 pCi/g compared to the current RG of 0.113 pCi/g; for Th-232, the BTV would be established at 2.21 pCi/g compared to the current RG of 1.69 pCi/g; for U-235, the BTV would be established at 0.245 pCi/g compared to the current RG of 0.195 pCi/g. We have several concerns about this approach.

First, the BTVs calculated for each RBA are based on datasets that, in three of the five RBAs, combine data from shallow and deep soils representing multiple statistical populations (as summarized in Tables 6-3 and 6-4). The Upper Simultaneous Level (USL), the statistic proposed for calculation of many of the BTVs, is not advised when the dataset used to calculate a BTV represents multiple populations.

Second, this approach, in which the highest BTV calculated for any of the five RBAs would be applied sitewide, effectively combines data from multiple areas representing different statistical populations (as summarized in Section 6.3.2.1 and Tables 6-2). As noted above, the USL is not advised when a dataset represents multiple populations.

Third, given the relatively large variability across the five RBAs, and the expected variability in the soils to be retested, applying the highest of the five BTVs sitewide is likely to overestimate the background radionuclide concentrations in an unknown but potentially significant fraction of the soils to be retested (i.e., “false negatives”).

An alternative approach would be to calculate BTVs based on a subset of the RBAs. We considered the use of a dataset that combined sample results from a subset of RBAs, as well as the use of data from a single RBA. We recommend the calculation of BTVs using only the offsite RBA-S data (San Bruno Mountain State and County Park). Data from the four onsite RBAs would not be used to calculate the BTVs. This approach would result in one BTV above its current RG, establishing the BTV for Cs-137 at 0.141 pCi/g compared to the RG at 0.113 pCi/g.

We support this alternative approach because, unlike most of the onsite RBAs, RBA-S appears to represent a single statistical population for five of the six radionuclides. This fulfills one of the requirements for use of the USL. This approach would make it unnecessary to combine data from multiple RBAs representing distinct statistical populations, does not increase an RG because of a BTV based on a detection limit (see comment

#14), and excludes the elevated 0.477 pCi/g Cs-137 result at RBA-4 (see comment #10).

This alternative approach would, compared to the Navy's preferred approach, decrease the chances of incorrectly concluding that soils are representative of background when they are actually affected by Site contamination (i.e., "false negatives") but increase the chance of incorrectly concluding that a sample result represents site-related contamination when it actually reflects background (a "false positive"). For several radionuclides, the highest concentrations measured in the four onsite RBAs exceed the BTV that would be calculated using RBA-S data alone. The frequency of false positives would be reduced but not eliminated by using the USL as the background statistic.

Given the potential for a non-negligible frequency of false positives, we would not necessarily conclude that a retesting result exceeding a BTV represents Site-related contamination. This consideration would apply to two of the radionuclides, Cs-137 and Th-232, where the RG would be equal to or close to the BTV. We propose that several criteria be used to determine whether a sample result that exceeds a BTV represents background or Site-related contamination, including: 1) whether the sample was collected in an area with a known or suspected release of the radionuclide; 2) whether the sample result exceeds the maximum concentration in soil with a similar soil type, color, and/or local environment; and 3) whether nearby samples also exceed the BTV.

For the other four radionuclides examined in the background study (Pu-239, Ra-226, Sr-90, and U-235), the RG exceeds the BTV by a greater margin compared to Cs-137 and Th-232, reducing the chances that a false positive (i.e., a retesting value above the BTV that represents background) would have any practical consequences.

Whichever approach is used to develop and apply the BTVs, there will be a risk of both false positive and false negative errors when Site data are compared to the BTVs. We believe that the use of RBA-S data alone better balances the relative risk of the two types of error compared to the Navy's preferred approach.